Technical Report 1177

Evaluation and Refinement of a Screening Instrument for U.S. Army Recruiters: Noncommissioned Officer Leadership Skills Inventory

Kristen E. Horgen, U. Christean Kubisiak, Valentina Bruk-Lee, Patrick W. Connell, Lisa M. Penney, Walter C. Borman, Victoria L. Pace, and Elizabeth Lentz
Personnel Decisions Research Institutes, Inc.

Leonard A. White, and Mark C. Young
U.S. Army Research Institute

Stephen V. Bowles
U.S. Army Recruiting Command

March 2006

20060509048

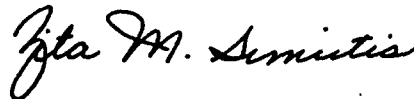United States Army Research Institute
for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS
Technical Director

ZITA M. SIMUTIS
Director

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) March 2006	2. REPORT TYPE Final	3. DATES COVERED (from... to) September 2003 – March 2006

4. TITLE AND SUBTITLE Evaluation and Refinement of a Screening Instrument for U.S. Army Recruiters: Noncommissioned Officer Leadership Skills Inventory	5a. CONTRACT OR GRANT NUMBER DASW01-03-D-0016 DO 0001
	5b. PROGRAM ELEMENT NUMBER 622785

6. AUTHOR(S) Kristen E. Horgen, U. Christean Kubisiak, Valentina Bruk-Lee, Patrick W. Connell, Lisa M. Penney, Walter C. Borman, Victoria L. Pace, & Elizabeth Lentz (Personnel Decisions Research Institutes, Inc.); Leonard A. White & Mark C. Young (U.S. Army Research Institute); and Stephen V. Bowles (U.S. Army Recruiting Command)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 255
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Personnel Decisions Research Institutes, Inc. 100 South Ashley Drive, Suite 375 Tampa, FL 33602	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-RS 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1177

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative and Subject Matter POC: Len White

14. ABSTRACT *(Maximum 200 words)*:
The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), the U.S. Army Recruiting Command (USAREC), and their contractor, Personnel Decisions Research Institutes, Inc. (PDRI) have collaborated to implement the online administration of the Noncommissioned Officer Leadership Skills Inventory (NLSI). The NLSI measures skills and abilities related to NCO performance, including work orientation, interpersonal skills, and leadership capability. We also conducted research to validate the NLSI as a predictor of U.S. Army recruiter performance. The NLSI was successfully implemented as the first online recruiter testing administered in proctored settings worldwide. The validation results indicate that the NLSI predicts recruiter training attrition and recruiters' duty performance, as measured by individual recruiter production (e.g., signed contracts).

15. SUBJECT TERMS
Army Recruiters, Noncommissioned Officer Leadership Skills Inventory, NLSI, Recruiter Production, Online testing

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	65	Ellen Kinzer Technical Publication Specialist 703/602-8047

Technical Report 1177

Evaluation and Refinement of a Screening Instrument for U.S. Army Recruiters: Noncommissioned Officer Leadership Skills Inventory

Kristen E. Horgen, U. Christean Kubisiak, Valentina Bruk-Lee, Patrick W. Connell, Lisa M. Penney, Walter C. Borman, Victoria L. Pace, and Elizabeth Lentz
Personnel Decisions Research Institutes, Inc.

Leonard A. White and Mark C. Young
U.S. Army Research Institute

Stephen V. Bowles
U.S. Army Recruiting Command

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

March 2006

Executive Summary

Research Requirement:

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and Personnel Decisions Research Institutes, Inc. (PDRI) have continued research to develop and refine a screening instrument to select Soldiers with high potential for success in recruiting duty. This measure, known as the Noncommissioned Officer Leadership Skills Inventory (NLSI), is a screening test battery that measures skills and abilities related to recruiter performance, including work orientation, interpersonal skills, and leadership capability. Although the instrument was validated previously in a concurrent context, PDRI was asked to assist in implementing online NLSI administration, and to examine the predictive validity of the NLSI against additional criterion measures. The overall objective was to eventually establish a screening process to identify Soldiers who are likely to perform successfully as recruiters and to select these Soldiers for recruiting duty prior to sending them to the recruiter training.

This report describes the successful first steps to implement online test administration for Soldiers assigned to recruiting duty, the development of several criterion measures of recruiter performance, and the results of the NLSI validation research.

Procedure:

The United States Army Recruiting Command, ARI and PDRI worked together with several Army agencies and private contractors to plan, test, and implement worldwide, online NLSI administration. The online version of the NLSI was administered to thousands of Soldiers around the world in 2003 and 2004.

PDRI also developed and/or collected several criterion measures of recruiter performance in training and on-the-job. We developed a criterion measure of individual recruiter production (i.e., average number of recruits enlisted per month) from United States Army Recruiting Command sources. In addition, we developed a multi-media rater training program and collected online performance ratings from recruiters and station commanders across the country.

These criterion measures were combined with NLSI data and background and demographic data to form the Predictive Validation Database. This database

contains information on almost 5,000 recruiters and can serve as the basis for future research on the NLSI and individual recruiter performance.

Finally, we used this information to refine the NLSI scoring key and analyzed the relationships between NLSI scores and various criterion measures, including training attrition from the Army Recruiter Course (ARC), measures of individual recruiter production, and peer and supervisor ratings of job performance.

Findings:

The results of the validation research demonstrate that the NLSI predicts both individual recruiter production and attrition from recruiter training. Recruiters with higher NLSI scores were more likely to graduate from recruiter training and had higher levels of individual recruiter production in the field. There were no significant mean differences in NLSI scores across race and gender groups, suggesting the use of the NLSI would not result in adverse impact. Other benefits, such as increased levels of job satisfaction, lower levels of stress, and higher quality of life may result from using the NLSI to select those Soldiers best suited for recruiting duty.

The validation data supports an initial use of the NLSI for screening a small percentage (e.g., 5%) of Soldiers who are a poor fit for recruiting duty. Ideally a large number of potential recruiter candidates can be screened, increasing the utility of the NLSI. The authors also recommend further testing of the NLSI for use as a classification tool for other Army NCO positions.

Utilization and Dissemination of Findings:

ARI and PDRI presented briefings and periodic updates to representatives of U.S. Army Accessions Command, U.S. Army Recruiting Command, Human Resource Command, and Army G-1 (i.e., briefing to MG Michael D. Rochelle, Commanding General, U.S. Army Recruiting Command, August 2003, as well as periodic updates to MG Rochelle during the entire course of the project; briefing to COL Jack Collins, Commandant Recruiting and Retention School, October 2004; briefing to BG Byrne, Army G-1, July, 2005; briefing to COL Norvel Dillard, Chief, Enlisted Accessions Division, September 2004). ARI and PDRI also presented this research at several professional conferences (e.g., Bowles et al., 2003; Borman et al., 2004). The final project briefing was presented to LTC Linda Ross, U.S. Army Recruiting Command Psychologist on April 29, 2005.

This research was intended to help the Army move forward with its future efforts to develop and implement a screening process for Army recruiters. The research

makes a significant contribution to understanding the determinants of Army recruiter job performance.

The NLSI and recruiter performance measures developed over the course of the research can be utilized for other purposes as well. For example, NLSI test scores were used to help evaluate students' progress during the Assessment Board at the Recruiting and Retention School. In addition, we developed several measures that can be used by USAREC for training and development purposes. The Recruiter Situational Judgment test can be used at the Recruiting and Retention School to help train new recruiters to effectively solve difficult recruiting situations. The Army Recruiter Performance Rating Scales can be used to as an assessment and development tool to review recruiters' performance on the job and specify areas for improvement. Finally, as a result of this research, there is a working system to deliver secure, proctored testing in Army DTFs around the world.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

The Department of the Army (DA) and the United States Army Recruiting Command (USAREC) must recruit more than 100,000 qualified young people each year for the Regular Army and the Army Reserve. Increasing economic growth and opportunity in civilian jobs, changes in educational aspirations among parents and their children, and negative perceptions of military life can make this task increasingly difficult (Kubisiak, et al., 2003). In addition, recruiters are now faced with the challenge of attracting new recruits during a time of prolonged conflict. Not only have these recent world events made recruiting more difficult, they have placed additional recruiting demands on USAREC. At the same time the Army is asked to recruit larger numbers of Soldiers, it is losing more Soldiers due to attrition (Goldberg, Kimko, & Lewis, 2005). USAREC has taken a number of steps to address these challenges, including increasing incentives, implementing new enlistment bonuses, and increasing the number of field recruiters.

The Army identifies and selects over 2,500 new recruiters each year from among their best Soldiers. These recruiters receive extensive training and work long hours in a demanding and stressful job. There are currently more than 7,300 Soldier and civilian recruiters in more than 1,600 recruiting stations throughout the U.S. and overseas (USAREC, 2005).

To further assist USAREC, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and Personnel Decisions Research Institutes, Inc. (PDRI) have conducted an initial concurrent validation of a screening instrument to select Soldiers with high potential for recruiting duty success. This measure, known as the Noncommissioned Officer Leadership Skills Inventory (NLSI), is a screening test battery intended to predict recruiter performance. Although the instrument was validated previously in a concurrent context, PDRI was asked to assist in implementing online NLSI administration and to examine the predictive validity of the NLSI against additional criteria. The tasks reviewed in this report are listed below.

Specifically, PDRI: (1) coordinated with a number of Army agencies and commercial vendors to implement worldwide, online NLSI administration; (2) tested recruiters on the NLSI and created the NLSI predictive validation database; (3) developed a criterion measure of individual recruiter production; (4) evaluated the feasibility of developing a measure of recruiter production that incorporated an index of recruit quality; (5) collected peer and supervisor performance ratings on recruiters in the predictive validation sample; (6) analyzed relationships between NLSI scores and various criterion measures, including

training attrition from the Army Recruiter Course (ARC), measures of individual recruiter production (i.e., average number of recruits enlisted per month), and performance ratings; and (7) refined the NLSI scoring algorithm developed from the concurrent validation based on the results from the predictive validation sample.

We briefly describe the background of the development of the NLSI and initial concurrent validation research in the next section. This is followed by a project overview and report outline.

Previous Research on Correlates of Recruiter Success

In August 1999, the Secretary of the Army tasked the Assistant Secretary of the Army for Manpower and Reserve Affairs and the U.S. Army Deputy Chief of Staff for Personnel to lead the Army in six major initiatives to eliminate recruiting shortfalls. One of the recruiting initiatives was to develop an immediate plan of action to improve the selection, training, equipping, and management of the recruiting force. As a result, a multi-year plan was developed to validate a new screening tool for selecting recruiters against measures of recruiter performance. The objective was to establish a screening process to identify Soldiers who are likely to perform successfully as recruiters, and to select these Soldiers for recruiting duty prior to sending them to the ARC.

Based on a review of recruiting research (Borman, Horgen, & Penney, 2000; Penney, Horgen, & Borman, 1999; Borman, Penney, et al., 2001) and inventories found to be successful for selection into military and civilian jobs similar to the Army recruiter job (Sutton, Horgen, Borman, & Kubisiak, 2001), we selected a battery of instruments to include in the concurrent validation research. Subsequently, ARI and PDRI conducted a concurrent validity in 2001 to evaluate the empirical validity of a paper-and-pencil battery for Army recruiters. In this concurrent validation research, several instruments effectively predicted recruiting success, as measured by ratings of recruiter performance and recruiter production (i.e., average number of recruits enlisted per month). These concurrent validation results are more fully described in Borman et al., (2003) and White, Borman, & Bowles (2001). Based on the promising results from the concurrent validation, ARI, USAREC, and PDRI began a large-scale effort to implement online NLSI testing worldwide and to investigate the predictive validity of these new instruments, collectively known as the NLSI, against several measures of recruiter performance in training and later sales success on the job.

Project Overview and Report Outline

The recruiter predictive validation project took place over several years and is briefly outlined below. The Recruiting and Retention School (RRS) at Fort Jackson began to administer the paper-and-pencil version of the predictor

measures in January of 2002. In January 2003, administration of the predictor was transitioned from a paper instrument to an online, computerized testing environment. Additional details regarding the data collection at the RRS and the transition to online test administration are provided in Chapter 2. Chapter 2 also describes the development and maintenance of the Recruiter Predictive Validation Database. The predictor instruments are described in Chapter 3.

As presented in Chapters 4 and 5, several measures of recruiter performance were developed and used to collect performance data during various stages of the project. These included measures of recruiter production from USAREC records, and both supervisory and peer ratings of job performance. In Chapter 6, relationships between the predictor and the various criterion measures are described. Finally, in Chapter 7, we summarize the results and make recommendations for future research.

Chapter 2 - Data Collection and Database Development

Approach

The research strategy chosen for this project was to obtain *criterion-related* evidence using a *predictive* validation design. We accomplished this by testing Soldiers before they reported to the Army Recruiter Course (ARC) for training, and then collecting measures of these same individuals' job performance at a later date. Test scores were then related to how well individuals performed on the job. Successful validation of this type provides additional evidence that the NLSI can be used to identify more qualified candidates. This validation methodology is one of three validation strategies presented in the *Uniform Guidelines on Employee Selection Procedures* (43 Federal Register 38290-38315, 1978), the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), and the Society for Industrial/Organizational Psychology's *Principles for the Validation and Use of Personnel Selection Procedures* (2003). The specifics of the data collection efforts, including development of the online NLSI, details of NLSI administration, and database development are provided in the following sections.

Participants & Procedures for Predictor Data Collection

NLSI data were collected from 4,586 recruiters from January 2002 through August 2004. From January 2002 through December 2003, 2,143 recruiters took the paper-and-pencil NLSI at the RRS during their first week at the ARC. Beginning in January 2003, another 2,443 recruiters were administered the online NLSI at a Digital Training Facility (DTF) at Fort Jackson, or at a DTF in any of 276 locations worldwide before beginning recruiter training at the RRS. Characteristics of the participants are detailed in Chapter 6. Below we describe the development and implementation of the online version of the NLSI and test administration procedures.

Development and Implementation of the Online NLSI

The original intention of the recruiter screening research program was to develop an instrument to screen Soldiers for potential assignment to recruiting duty *before* they were transferred from their Military Occupational Specialty (MOS) to the ARC for recruiter training. To achieve this objective, ARI, USAREC, and PDRI worked to transition the paper-and-pencil NLSI to a computerized version

administered at DTFs with the capability to test geographically diverse applicants in a secure, cost-effective manner. This transition from paper-and-pencil testing to the online NLSI represents a successful first attempt at implementing a proctored, online Army personnel testing system at the DTFs.

Soldiers assigned to recruiting duty were tasked by the Army Human Resources Command (HRC) to test at their local DTF before arriving at the RRS at Fort Jackson. Some students were unable to test prior to their arrival and were tested at the DTF located at Fort Jackson. When complete, test data were transmitted via secure, encrypted electronic connections between the Army Training Support Center (ATSC), the DTFs, HRC, ePredix, and PDRI. Ultimately, the NLSI is to be migrated to the Learning Management System (LMS), the Army's new software environment for computerized training delivery.

The content and instructions in the NLSI remained the same across the paper-and-pencil and online versions. To the extent possible, the items' on-screen appearance was kept similar to the paper version. For example, in the paper-and-pencil version of the NLSI, test-takers indicated their responses by filling in bubbles on answer sheets that were mechanically scanned. For the online version, test-takers indicated their responses by clicking a mouse.

The online format allowed PDRI to make several improvements in the testing procedures. With regard to skipping items, which cannot be controlled on a paper-and-pencil test, we determined that users should be required to respond to each item before going on to the next item. This eliminated analytical complications resulting from missing data. To further maintain similarity to the paper-and-pencil format, test-takers were given the capability to move back and forth between screens and change responses to items if they so desired. Testing time was limited to three hours for the whole NLSI, but test-takers could take as much time as needed on each item. If they had to stop testing, they were allowed to resume from where they left off, provided they returned to finish within 30 days.

The transition to online administration of the NLSI required that we address several administrative issues. The first online version of the NLSI, the one used for the work described in this technical report, was hosted by ePredix, Inc., an organization that specializes in online testing. Implementation, score reporting, and data sharing required months of coordination between the ATSC, the Army HRC, the Distributed Learning System (DLS) group, USAREC, ARI, PDRI, and ePredix. The NLSI was only administered at the Army's DTFs, where DTF proctors were able to securely connect test-takers to ePredix's hosting system. This provided a host of benefits in that 276 DTF facilities are available at Army bases throughout the world and are similarly equipped and operated. Therefore, the technical requirements of the web-site could be tailored to the Army's technology platform. Further, the security of the test content could be maintained by allowing access only from the DTFs.

Another benefit to using DTFs for test administration was that proctors were able to monitor the testing. This had significant benefits for the NLSI in a number of ways. First, proctors checked the identification of the Soldiers presenting themselves for testing. Second, proctors monitored testing according to Army policies and procedures, enhancing the uniformity of the testing across locations. Third, proctors were available to provide technical support and answer questions as needed. Finally, proctors were required to log authorized users into the system, further enhancing the security of the test.

Online NLSI administration also streamlined the score reporting procedures. Scores were computed immediately after the test-takers completed the NLSI and reported back to Army decision makers at the RRS and HRC. NLSI test scores were reported for use in the Assessment Board at the RRS. In the Assessment Board process, the Command Psychologist and the ARC instructors used NLSI overall and scale scores, along with student performance and instructor evaluations, to evaluate students' progress in the ARC. Scores were also reported to PDRI to populate a database for test validation research purposes.

Overall, the development and implementation of the online NLSI allowed for the secure and cost-effective testing of thousands of Army recruiter candidates worldwide. Test scoring and reporting was streamlined and available for immediate use, whether for administrative or research purposes.

Test Administration Procedures

Both the paper-and-pencil and online NLSI were administered in proctored settings. USAREC or DTF proctors were trained in testing procedures and administered the test. USAREC, ARI, and PDRI developed protocols and trained proctors to maintain standardized test administration procedures across facilities.

The NLSI was administered under high-stakes testing conditions (as opposed to for-research-only conditions). Examinees read a special set of instructions before beginning the NLSI. The instructions informed examinees that Army decision-makers may use the results to make future assignments. Soldiers were also instructed that the test was designed to detect deliberate attempts to collaborate with others in answering the items. In addition, participants were asked to read a Privacy Act statement. The NLSI took approximately 1.5 hours to complete.

Database Development

The NLSI Predictive Validation Database drew from several data sources. The database grew as data elements were added from new sources, as new participants were tested on the NLSI, and as we conducted validity analyses on an on-going basis over the two years of the project. Currently, the database consists of 4,998 participants and 394 variables.

In addition to NLSI raw and computed scores, archival data from several Army sources were added to the database throughout the project. In particular: (1) demographic information (e.g., gender, race), basic military background data (e.g., basic active service date), and Armed Services Vocational Aptitude Battery (ASVAB) scores were obtained from the Total Army Personnel Data Base (TAPDB) and Defense Manpower Data Center (DMDC) data files; (2) ARC attrition data were obtained from the Army's Training Requirements and Resources System (ATRRS); (3) ARC performance data (e.g., test scores, instructor evaluations) were obtained from USAREC; (4) recruiter detail relief data (e.g., relief from recruiting duty) were obtained from USAREC; (5) recruiter production data (e.g., number and type of recruits enlisted each month) were obtained from USAREC. In addition to the archival data, performance rating data were collected by PDRI.

Chapter 3 – Noncommissioned Officer Leadership Skills Inventory (NLSI)

Part I

Part I of the NLSI is a 125-item self-report questionnaire that measures prior behaviors and reactions to specific life events that are indicative of such areas as leadership, interpersonal skills, and integrity. Previous research has demonstrated that the scales predict delinquency criteria, Special Forces field performance, completion of the Special Forces Assessment and Selection course, and disciplinary infractions among NCOs and first term enlisted personnel (e.g., Kilcullen, Chen, Zazanis, Carpenter, & Goodwin, 1999; Kilcullen, Mael, Goodwin, & Zazanis, 1999). Additionally, in research with Army civilians, the Tolerance for Ambiguity, Openness, Emergent Leadership, and Social Perceptiveness scales were related to effective job performance (Kilcullen, White, Zacarro, & Parker, 2000). Thus, the NLSI Part I has demonstrated evidence for criterion-related validity in military and non-military settings. Moreover, in specific relation to recruiter performance, the scales assess constructs that are theoretically relevant to success in recruitment such as Social Perceptiveness and Interpersonal Skills (Sutton, et al., 2001). The Part I scales and definitions can be found in Appendix A.

Part II

Part II of the NLSI is a 34-item self-descriptive inventory. The scales assess personality-like traits relevant to military performance including Work Motivation, Agreeableness, Dependability, and Dominance. Part II scales and definitions can be found in Appendix B. Each item consists of four behavioral statements that represent different personality constructs. Within each tetrad, examinees are asked to select one statement that is most like them and a different statement that is least like them.

In a series of investigations, the scales used in Part II have been shown to predict Soldier attrition and performance during the first term of enlistment (Young, Heggestad, Rumsey, & White, 2000; Young, McCloy, Waters, & White, 2004; Young, White, Heggestad, & Barnes, 2004). In addition, preliminary findings indicate that these measures are more resistant to faking than other instruments, such as the Assessment of Background and Life Experiences (ABLE; Young et al.; White & Young, 2001). In other research, several of these scales were linked to Special Forces field performance (Kilcullen, Chen, et al., 1999), Correctional

Specialist performance, and the successful completion of Explosive Ordnance Disposal (EOD) training (White & Young, 2001). These results suggest that this instrument has promise for measuring constructs important for Soldier job performance, and several scales measure constructs theoretically relevant for successful recruiter performance (e.g., Work Motivation, Adjustment).

Part III – Situational Judgment Test

As part of the concurrent validation of the recruiter screening tool, an Army Recruiter Situational Judgment Test (AR-SJT) was developed as a criterion measure of recruiter performance. The AR-SJT presents a series of 25 challenging, but realistic, situations that recruiters might encounter on their job and asks the test-taker to indicate which of four response options he/she believes is the best way to handle the situation. Responses are scored by comparing them to recruiter expert judgments of the effectiveness of each response. The SJT development process is described in Borman, Horgen, et al. (2001).

A revised SJT was developed as Part III of the NLSI. However, using the SJT for predictive purposes, rather than as a criterion, required a new approach to item development. The items on the AR-SJT contain a great deal of detail regarding specific recruiting tactics, procedures, and regulations (Horgen, Penney, Birkeland, Kubisiak, & Borman, 2001). Items with this job-specific recruiting content would not be appropriate for Soldiers with no recruiting experience or training. However, we wanted to retain the sales content of the AR-SJT to try to identify those Soldiers with the interpersonal skills and abilities that might make them successful recruiters.

After examining the content of the items, we determined that many could be rewritten to remove the recruiting content and still retain the sales aspects of the items. A subset of the items was rewritten to carefully preserve the theme of the original recruiting situation. For example, a situation about making an Army recruiting presentation to a recruit was changed to a situation about making a sales presentation to a potential client. The response options were similarly changed so that behaviors in the original recruiting-specific item remained virtually the same. For example, a response option, 'get help from another recruiter in the recruiting station' was changed to 'get help from another salesperson in the office'. Several items were too specific to Army recruiting to be rewritten appropriately, and these items were not included in the revised version.

In addition to these sales items, other items with leadership content were added, based on their success in past Army research with junior Noncommissioned Officers (Borman, et al., 1990; Hanson & Borman, 1995). These describe situations that second tour Soldiers might encounter and were intended to apply to Soldiers in any MOS.

Chapter 4 – Development of Criterion Measures of Recruiter Production

Recruiter Production

A primary measure for evaluating Army recruiter effectiveness is an index of individual recruiter productivity. Typically, such measures focus on the number of recruits or contracts signed in a specific time period. Productivity measures are important to the Army because USAREC's mission is to recruit Soldiers for the Army, and it is held accountable for accessing a certain number of new Soldiers each year. While USAREC maintains a database of recruiter production, this database is structured in such a way that it required substantial effort to develop an individual monthly production index for each recruiter in our predictive validation database (Penney, 2004).

Individual Production Average for the Validation Research

USAREC tracks monthly production information for all Army recruiters. As part of the validation process, participants were followed from their initial NLSI testing through up to 28 months of recruiter service. For the current project, we obtained monthly information from January 2002 through April 2004 regarding the gross number of recruits signed, the number of recruits that dropped out of the Delayed Entry Program (DEP loss), and the net number of recruits signed (gross production minus DEP loss) for every recruiter in USAREC. Descriptive statistics for production in USAREC and the validation sample are presented in Tables 1 and 2.

Table 1. Raw Production: Descriptive Statistics for All of USAREC (N = 13,307)

	Minimum	Maximum	Mean	Standard Deviation
Average Monthly Gross Production	0	7.89	0.99	0.67
Average Monthly Net Production	-2.00	7.89	0.82	0.62

Note: A negative net production value indicates that a recruiter lost a recruit or recruits during DEP and this recruit did not access into the Army.

Table 2. Production: Descriptive Statistics for Validation Sample (N = 2,883)

	Minimum	Maximum	Mean	Standard Deviation
Average Monthly Gross Production	0	5.23	1.24	0.54
Average Monthly Net Production	-1.00	5.00	1.09	0.52

Note: A negative net production value indicates that a recruiter lost a recruit or recruits during DEP and this recruit did not access into the Army.

Recruiters who participated in this research were tested prior to their arrival or during their first week at the RRS. Therefore, the number of months recruiters have been in the field varies from one to 28 (see Table 3). Although data were obtained for a 28-month period, not all recruiters had data available for all 28 months. On average, recruiters in the current sample had 13.86 months of production data.

Table 3. Total Number of Months Production Data Available for Recruiters in the Predictive Validation Sample (N = 2,883)

Number of Months Production Data Available	N	Number of Months Production Data Available	N
28	2	14	183
27	1	13	115
26	10	12	64
25	14	11	150
24	52	10	197
23	85	9	29
22	231	8	42
21	190	7	116
20	169	6	32
19	170	5	66
18	146	4	252
17	128	3	39
16	125	2	11
15	128	1	136

The production average scores calculated for use in this research were determined by taking the mean of the contracts signed per month by individual recruiters. In other words, for each recruiter, the total number of contracts signed between January 2002 and April 2004 was divided by the total number of months that the recruiter was actively recruiting. The creation of this production index presented a

considerable challenge. Unlike USAREC's write rate, which is calculated by obtaining a monthly average of production across recruiters, our index was based on obtaining a production average for each recruiter across time.

In order to calculate a production average at the individual level, PDRI reformatted the production data in USAREC's database for all 13,307 recruiters. This required a number of transformations to the data using both SAS and SPSS software. Ultimately, we created a database in which each recruiter had one line of data with up to 28 months of production, thus allowing us to calculate individual level averages.

To determine the total number of months each recruiter was on-production, we used the month associated with recruiters' first appearance in USAREC's production database as their first month in the field and counted the recruiter as on-production in every month subsequent to that. For example, if the month associated with a recruiter's first appearance in the database was March 2002, that recruiter was considered to be on-production from March 2002-April 2004 for a total of 25 months.

As stated previously, the individual production average for some recruiters was based on as few as 3-4 months of data, whereas others had as many as 28 months. Because the stability of the production average is likely to be higher when more months of data are averaged, including averages based on only a few months of data may attenuate the observed relationships with the predictor and other criteria. Therefore, we examined the reliability of production averages based on varying number of months' data (see Table 4) to determine an appropriate cut-off. Based on these findings, as well as a concern for retaining a large sample, we decided to screen out those individuals with less than four months of production data.

Table 4. Reliabilities of Production Indices Using Different Time Intervals

Time Length	Reliability
12 months	.72
11 months	.62
10 months	.79
9 months	.72
8 months	.64
7 months	.55
6 months	.44
5 months	.42
4 months	.41
3 months	.39
2 months	.27

Based on our prior work with production data in the concurrent validation (Borman et al., 2003; White et al., 2001), we determined that the number of contracts signed, or gross production, was a more appropriate measure than the number of accessions or net production. The previous research found that gross and net production were highly correlated ($r = .97$). They were also highly correlated in the predictive validation sample ($r = .97$). However, in the concurrent validation, analyses indicated that gross production was more reliable over a 12-month period than net production (Spearman corrected $r = .68$ for gross and .57 for net). One possible explanation for this finding is that factors beyond the control of recruiters may account for significant variance in DEP attrition, more so than for gross production itself. Therefore, we decided that gross production would be a more reliable indicator of recruiter effectiveness than net production.

In the concurrent validation, the production data were adjusted to account for differences in recruiting difficulty across months and locations around the country (Penney, Horgen, Kubisiak, Borman, & Birkeland, 2002). However, these corrected individual production averages were very highly correlated with the raw production averages ($r = .98$). Therefore, to simplify interpretation of the data, only raw production averages were used in this research. The gross production monthly average index for individual recruiters was used in subsequent validation analyses, correlating production both with other performance criteria (e.g., peer and supervisor ratings) and, most importantly, with predictor test scores.

We also attempted to create an individual recruiter production index to account for recruit quality. As these findings are not central to the validation research, the information regarding this work is presented in Appendix C.

Chapter 5 – Development of a Criterion Measure of Online Recruiter Job Performance Ratings

Results from previous analyses investigating the concurrent validity of some of the measures included in the NLSI against performance ratings were encouraging (White et al., 2002). Thus, we attempted to replicate those findings in a predictive context. However, peer and supervisor ratings were more difficult to collect in this context, as recruiters' peers and supervisors were located in recruiting stations across the United States. The goal of this portion of the project was to develop a method to collect peer and supervisor ratings of recruiter performance from a group of peer recruiters and station commanders located across the country. Thus, an online version of the recruiter job performance rating scales was created.

Army Recruiter Performance Rating Scales

Behavior-based rating scales were used to measure the job performance of recruiters. A previous report describes development of the recruiter performance rating scales (Borman, Horgen, et al., 2001). These same scales were used successfully in the concurrent validation research (Penney, et al., 2002). The behavior-based rating scales were designed to encourage raters to make evaluations as objectively as possible. Specifically, within each performance dimension, statements describing behaviors that reflect performance that is "very effective", "effective", "needs some improvement", and "needs considerable improvement" anchor these four effectiveness levels on the scales. Raters were asked to compare observed recruiter behavior with the statements that anchor the different effectiveness levels on each dimension. The Army Recruiter Performance Rating Scales appear in Appendix D.

The eight behavioral dimensions are: (1) Locating and Contacting Qualified Prospects; (2) Gaining and Maintaining Rapport; (3) Obtaining Information from Prospects and Making Good Person-Army Fits; (4) Salesmanship Skills; (5) Delayed Entry Program (DEP)/Delayed Training Program (DTP) Maintenance; (6) Establishing and Maintaining Good Relationships in the Community; (7) Organizing Skills/Time Management; and (8) Supporting Other Recruiters and USAREC.

The first four dimensions clearly represent the major steps that recruiters perform in the applicant contracting process. DEP/DTP Maintenance was identified to capture the Army's interest in sustaining relationships with new recruits through the delayed entry process. Recruiters must also initiate, develop, and maintain

productive relationships with individuals and agencies in the community in order to build and enhance the Army's reputation. Planning, organizing, and time management skills refer to the recruiter's ability to balance priorities and deadlines and to manage enlistment processing. The final behavioral dimension includes coordinating with and supporting other recruiters, following orders, and helping or mentoring other recruiters. Taken together, these eight dimensions define the Army recruiters' behavioral performance requirements.

Online Criterion Data Collection

We developed a plan to collect performance ratings from recruiters' peers and supervisors. Discussions with ARI and USAREC indicated that both sources should provide valuable information regarding recruiters' performance. Also, obtaining ratings from multiple raters for each ratee increases the interrater reliability of the ratings.

In the context of the NLSI validation, the traditional approach to gathering ratings had to be modified to meet a number of challenges. Because of the number of raters and their geographic dispersion, it was impractical to travel to their locations and provide rater training. Asking the raters to travel to centralized locations would have been prohibitively expensive and time consuming, as well. Additionally, recruiters and station commander supervisors have many demands placed on their time and the rating task had to be minimally intrusive. Further, given the limited time that the NLSI has been in use, obtaining the number of ratings required for the validation required a very high response rate.

Recognizing these constraints, PDRI developed an innovative solution to collect ratings online. This allowed raters to make their ratings anywhere they had internet access. Because Army recruiters are issued notebook PCs, we could be certain that the raters had internet access, were able to use computers well enough to navigate the web-site, and that the PCs themselves were sufficiently standardized that they would run the software.

Online Rater Training

The online rating format also provided us with an opportunity to try a new approach to rater training. Because the in-person component of the training could not be used, we developed a seven minute, CD-ROM-based, multi-media rater training program. The presentation consisted of a virtual trainer verbally presenting instructions, written instructions that could be read at the viewer's own pace, and a walk-through of the rating task, demonstrating actual screens that the rater would see on the web-site.

The rater training program was designed to accomplish the following objectives: (1) orient raters to the rating task; (2) familiarize raters with the performance dimensions and how each is defined; (3) train raters to match observed recruiter

behavior with the behavioral summary statements to obtain a rating for each dimension; (4) describe common rater errors (e.g., halo); and (5) encourage raters to be as accurate as possible when making their ratings.

After the instructional video concluded, the software provided a link to the web-site designed for the completion of performance ratings. By requiring that the software presentation be played before allowing access to the web-site, we could ensure that the user was presented with the rater training. It also provided a seamless link from the rater training to the actual rating task, increasing the likelihood that the raters would stay with the task all the way through completion.

ARI/PDRI mailed the CD-ROM to the station commanders and recruiters in the participants' recruiting station. Both the supervisors (station commanders) and a subset of peers (recruiters) rated the participants. Included with the CD-ROM was a cover letter describing the NLSI validation effort, the purpose of the rater's involvement, an identification number and password for the raters, and identification numbers and names of the recruiters they would be rating. This letter also included contact information for the project team members so that raters could contact us if they encountered any technical problems or wished to ask questions. Raters were also instructed that their ratings would be used for research-purposes only and would be kept confidential.

After viewing the rater training and upon reaching the web site, raters were asked to provide their identification number and password. The next screen requested basic demographic information, and raters were asked to select from a list the name of the recruiter they would be rating. This methodology ensured that the raters were the intended individuals, and that they were rating recruiters with whom they did, in fact, work. Raters were also asked to indicate how long they had worked with the ratee(s). After completing their ratings, the users could close the application in completion of the task.

Approximately two weeks after the materials had been mailed out to the raters, project staff called individuals who had not yet responded and asked them to complete their ratings. This methodology yielded a 70 percent response rate for the raters, much higher than what can typically be accomplished using a traditional mail-out approach.

Utilizing the online methodology provided a number of additional advantages over the traditional paper-and-pencil, in-person method. For example, it was less costly and time consuming to create and send the CD-ROMs to the raters than it would have been to have project staff travel to many different locations. Further, the demand on the raters' time was kept to 20 minutes to complete the entire task, including training and making the ratings. Additionally, the multi-media presentation maximized uniformity of training across locations. Finally, the online data system allowed us to track, in real time, who had responded and completed their ratings. This allowed for a targeted follow-up of raters who had not completed the task.

Performance Rating Responses

In total, we contacted 647 raters to provide performance information on a subset of the recruiters in our sample. Table 5 shows the number of participating raters from each brigade and the total number of ratings collected.

Table 5. Total Number of Raters and Ratings from Each Brigade

Brigade	Raters	Ratings
1st Brigade	123	162
2nd Brigade	88	108
3rd Brigade	65	75
5th Brigade	96	126
6th Brigade	146	199

In total, performance ratings for 388 recruiters were collected from 304 peer and 219 supervisor raters. Individual raters were removed from the sample if they met at least one of two criteria. First, if the information provided by an individual rater appeared inaccurate (e.g., if the same rating was given to a recruiter across all eight dimensions), the rater was dropped. Based on this criterion, a total of five rater-ratee pairs were eliminated from the sample. In addition, we asked raters how long they had worked with the recruiter(s) they were evaluating. Interviews with recruiters indicated that raters who had worked with recruiters for less than 4 months likely had insufficient time to observe and accurately evaluate their performance. Based on this criterion, 19 additional rater-ratee pairs were eliminated from the data set. As a whole, the mean number of months raters had worked with recruiters was 11.34 months for peer raters and 9.69 months for supervisor raters.

The final sample included performance ratings for 380 recruiters. Ratings were provided by 300 peers and 206 supervisors. Table 6 shows the number of supervisor and peer raters for each recruiter.

Table 6. Number of Supervisor and Peer Raters

Number of Supervisor Raters per Ratee	N	Number of Peer Raters per Ratee	N	Total Number of Raters per Ratee	N
1	237	1	197	1	179
2	6	2	65	2	150
3	1	3	17	3	41
		4	4	4	6
				5	4

Mean number of supervisor raters per ratee = 1.03
Mean number of peer raters per ratee = 1.39
Mean total number of raters per ratee = 1.70
Note: Some raters served as both supervisor and peer raters for different recruiters.

Tables 7 and 8 illustrate the distribution of ratings across the 10-point rating scale for supervisor and peer raters. There is a low, but noteworthy percentage of ratings at the lower, ineffective end of the scale for both the peer and supervisor ratings. Many of the ratings fall in the 6-7 range, but overall, there is reasonable variability in both sets of ratings, suggesting that both supervisor and peer raters were differentiating between the more and less effective recruiters. Means and standard deviations across all the ratings were: 6.40 and 1.47 for supervisor raters, 6.42 and 1.43 for peer raters.

Table 7. Number and Percentage of Supervisor Ratings at Each Scale Point

Rating Scale Point (1=Lowest 10=Highest)	Number of Ratings[a]	Percentage of Ratings
1	20	1
2	49	3
3	89	5
4	162	8
5	269	14
6	390	20
7	381	20
8	259	13
9	220	11
10	88	5

[a]Total number of supervisor ratings across all eight dimensions.

Table 8. Number and Percentage of Peer Ratings at Each Scale Point

Rating Scale Point (1=Lowest 10=Highest)	Number of Ratings[a]	Percentage of Ratings
1	35	2
2	50	2
3	77	4
4	149	8
5	223	12
6	391	20
7	392	20
8	319	17
9	207	11
10	71	4

[a]Total number of peer ratings across all 8 dimensions.

Table 9 presents the reliabilities for the supervisor and peer ratings combined. In general, the reliabilities are quite high. Both rating sources provide important performance information because of their unique perspectives; and the higher reliabilities for both sources taken together supports the use of an aggregated supervisor/peer rating criterion.

Table 9. Interrater Reliabilities for Combined Supervisor, Peer Ratings[a]

Rating Dimension	Combined Peer/Supervisor Reliabilities[b]
Locating and Contacting Qualified Prospects	.65
Gaining and Maintaining Rapport	.55
Obtaining Information From Prospects and Making Good Person-Army Fits	.45
Salesmanship Skills	.63
DEP/DTP Maintenance	.51
Establishing and Maintaining Good Relationships in the Community	.44
Organizing Skills/Time Management	.62
Supporting Other Recruiters and USAREC	.54
Rating Composite	.70

[a]Reliabilities are intraclass correlation coefficients (ICC 1,k; Shrout & Fleiss, 1979).
[b]$N = 646$, k (harmonic mean) = 1.41

Rating scores were created for each recruiter by calculating the mean peer rating and the mean supervisor rating, and then averaging these two for each dimension. Table 10 shows the means and standard deviations of these combined rating scores for each dimension.

Table 10. Mean and Standard Deviations for Mean Ratings on Each Dimension

Rating Dimension	Mean[a]	Standard Deviation
Locating and Contacting Qualified Prospects	6.03	1.92
Gaining and Maintaining Rapport	6.95	1.82
Obtaining Information From Prospects and Making Good Person-Army Fits	6.12	1.70
Salesmanship Skills	6.07	1.87
DEP/DTP Maintenance	6.96	1.75
Establishing and Maintaining Good Relationships in the Community	6.34	1.80
Organizing Skills/Time Management	5.84	1.95
Supporting Other Recruiters and USAREC	6.57	1.89

[a]N = 380

Factor Analysis of the Ratings

To examine the underlying structure of the eight rating scale dimensions, we conducted a principal factors analysis with a varimax rotation on the combined supervisor/peer dimensional ratings. Results of these analyses suggest that a one-factor solution is the most interpretable description of the data. Table 11 shows the results of this factor analysis.

Table 11. Factor Loadings for Each Rating Dimension

Rating Dimension	Factor Loadings
Locating and Contacting Qualified Prospects	.86
Gaining and Maintaining Rapport	.75
Obtaining Information From Prospects and Making Good Person-Army Fits	.80
Salesmanship Skills	.78
DEP/DTP Maintenance	.65
Establishing and Maintaining Good Relationships in the Community	.69
Organizing Skills/Time Management	.67
Supporting Other Recruiters and USAREC	.72

Correlations between Criterion Measures

We investigated relationships between each of the eight rating scale dimensions and recruiter production. These results are displayed in Table 12.

Table 12. Correlations between Combined Peer/Supervisor Ratings and Recruiter Production

Rating Dimension	Gross Production Average
Locating and Contacting Qualified Prospects	.34*
Gaining and Maintaining Rapport	.27*
Obtaining Information From Prospects and Making Good Person-Army Fits	.27*
Salesmanship Skills	.39*
DEP/DTP Maintenance	.17*
Establishing and Maintaining Good Relationships in the Community	.28*
Organizing Skills/Time Management	.20*
Supporting Other Recruiters and USAREC	.22*
Rating Composite	.34*

*Correlation is significant at the 0.01 level (2-tailed)

As can be seen in Table 12, all eight rating scale dimensions as well as the ratings scale composite correlated significantly with recruiter production. Thus, in comparison to recruiters with a lower monthly production average, those with a higher monthly production average were rated more favorably by both their peers and supervisors. Results also showed that the rating scale dimensions with the closest theoretical connection to recruiter production (i.e., Locating and Contacting Qualified Prospects, Salesmanship Skills) showed the strongest relationships with the production index.

As a whole, these results support the use of an aggregated supervisor/peer rating criterion, in addition to the individual production criterion measure, in the predictive validation of the NLSI. Results of the validity analyses are presented in Chapter 6.

Chapter 6 - Validation Results

This chapter describes the sample used for the predictive validation analyses and summarizes the relationships found between the NLSI and measures of individual recruiter production and supervisory and peer ratings of recruiters' job effectiveness. In addition, relationships were examined between the NLSI and recruiters' success in training (vs. attrition) and recruiting duty relief after training.

Recruiter Sample

Table 13 provides demographic information for all the recruiters who completed the NLSI from January 2002 through August 2004. This information was obtained from Army databases.

Table 13. Composition of Total Sample			
Category		Frequency	Percent
Gender	Male	3,544	91.4
	Female	335	8.6
	Missing information	707	
	Totals	4,586	100.00
Race/Ethnicity	Black	973	25.6
	Caucasian	2,267	59.6
	Hispanic	414	10.9
	American Indian/Alaskan Native	19	0.5
	Asian/Pacific Islander	93	2.4
	Other/Unknown	40	1.1
	Missing information	780	
	Totals	4,586	100.00

Table 13. Composition of Total Sample (continued)			
Age	20 - 25	407	10.7
	26 - 30	1,134	29.8
	31 – 35	1,395	36.7
	36 – 40	721	18.9
	40+	149	3.9
	Missing information	780	
	Totals	4,586	100.00
Pay grade	E1	2	0.1
	E2	21	0.6
	E3	67	1.8
	E4	438	11.5
	E5	1,674	44.0
	E6	1,457	38.3
	E7	144	3.8
	E8	1	0.0
	Missing information	780	
	Totals	4586	100.00

Predictive Validation Sample

Of the approximately 4,500 recruiters who completed the NLSI, 2,860 had at least 4 months of production data available. Table 14 presents the demographic information for the sample used in the predictive validation analyses. This sample has very similar demographics compared to the total sample of recruiters in Table 15 above.

Table 14. Composition of Predictive Validation Sample			
	Category	**Frequency**	**Percent**
Gender	Male	2,582	90.2
	Female	242	9.8
	Missing information	36	
	Totals	2,860	100.00

Table 14. Composition of Predictive Validation Sample (continued)			
Race/Ethnicity	Black	733	26.7
	Caucasian	1615	58.7
	Hispanic	289	10.5
	American Indian/Alaskan Native	15	0.5
	Asian/Pacific Islander	67	2.4
	Other/Unknown	31	1.1
	Missing information	110	
	Totals	2,860	100.00
Age	20 - 25	142	5.1
	26 - 30	797	29.0
	31 – 35	1,037	37.8
	36 – 40	635	23.1
	40+	134	4.8
	Missing information	115	
	Totals	2,860	100.00
Pay grade	E1	1	0.0
	E2	5	0.2
	E3	42	1.5
	E4	228	8.3
	E5	1,203	43.7
	E6	1,146	41.7
	E7	123	4.5
	E8	0	0.0
	Missing information	112	
	Totals	2,860	100.00

Development of a New Empirical Key for Parts I and II

As described in Chapter 1, a concurrent validation (CV) was conducted and a rational-empirical key for the NLSI was developed. Although this CV key demonstrated moderate correlations with the criterion measures in the concurrent validation it was hypothesized that the validity of the scoring key could be refined, and possibly improved, by using the longitudinal production data and attrition criteria available in the large sample predictive validation. In addition, several new items were added to the NSLI administered in the predictive

validation that might add to its criterion-related validity. The following section describes the development of a new empirical key for Parts I and II of the NLSI.

Recruiter Production and Performance Ratings

The sample was split into development and cross-validation sub-samples. PDRI used the development sub-sample to derive an empirical key to predict the criterion measures. Item selection was guided by item-level correlations with production, performance ratings, and attrition from recruiter training in the development sample. Sixty-four items were selected from both Parts I and II for the new NLSI empirical key. The 64-item NLSI had an internal consistency reliability estimate of .90. Table 15 shows both the uncorrected and corrected correlations between the new NLSI final scoring key and the criterion measures in the development and cross-validation samples. The new empirically-keyed NLSI was significantly correlated with recruiter production. NLSI scores were not significantly related to performance ratings. However, the cross-validation sample size for performance ratings was quite small, and we gave preference to items that correlated most highly with production in the item selection process.

Table 15. Correlations Between the New NLSI Empirical Key and Criterion Measures

Empirically-keyed NLSI Total Score	Production		Performance Ratings	
	Uncorrected	Corrected	Uncorrected	Corrected
Development sample[a]	.19*	.22*	.35*	.42*
Cross-validation sample[b]	.16*	.19*	.06	.07

Note: Correlations were corrected for criterion unreliability.
[a] N = 2894 for production, N = 377 for performance ratings
[b] N = 908 for production, N = 118 for performance ratings
*$p < .01$

Figure 1 illustrates the relationship between NLSI scores and recruiter production in a sample of 2,860 recruiters with four or more months of production data. We dropped cases with less than four months of data to achieve an adequate degree of reliability in the production criterion measures (see Chapter 4). Recruiters scoring in the lowest 5% of cases on the NLSI have lower production averages than those with higher NLSI scores (1.02 vs. 1.21 contracts and 0.88 vs. 1.05 accessions per month).
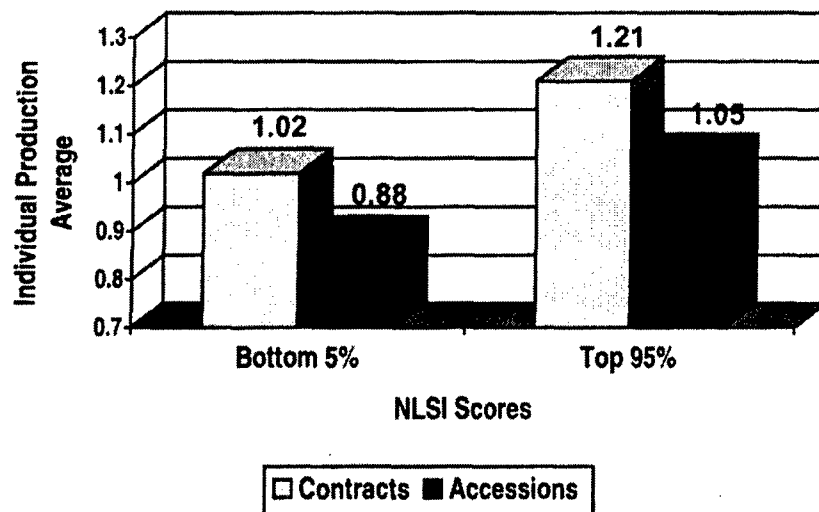
Figure 1. NLSI Scores Predict Recruiter Production.

We also investigated the performance of the CV key in the predictive validation sample. In a comparison between the items in the CV key and the new empirical key based on the predictive validation, we found that many of the items on the CV key continued to work as predictors of recruiter success. In fact, in the predictive validation sample, recruiter's NLSI score based on the CV and PV keys were highly correlated (r=.83).

The NLSI key had a cross-validity of .22 (N = 294) against recruiter production in the CV sample. As expected, this correlation showed some shrinkage in the PV sample where the uncorrected cross-validities using the CV key were .14 (N = 863) for the production criterion, and -.01 (N = 118) for the performance ratings criterion. By comparison, the cross-validity of the key developed from the PV sample was .16 against individual production (see Table 17), which was slightly higher than the uncorrected validity of .14 for the original CV key. Thus, some of the validity of the original CV key was restored by developing a new key using the PV sample data.

In addition to the total NLSI score, we examined the relationships between the NLSI scales and the production criterion. The individual NLSI scales had uncorrected correlations with recruiter production ranging from .01 to .13, with a median correlation of .07 (Ns range from 760-2,809). In sum, the validation results for production are encouraging, as the NLSI appears to be consistently related to recruiter production in two different samples through a variety of recruiting environments, in both concurrent and predictive research.

Race and Gender Analyses

Demographic data were available for a subset of cases in the predictive sample. Table 16 shows mean NLSI scores by gender and race/ethnicity. There was no significant difference between males and females on the NLSI. Ethnic group comparisons were made between Caucasians, Blacks, and Hispanics. Other ethnic groups were not included in the analyses because of the small sample sizes associated with these groups. Comparisons indicate that Hispanics scored significantly higher on the NLSI than both Blacks and Caucasians ($p < .05$). There were no significant differences between Caucasians and Blacks on the NLSI. These preliminary findings suggest that the NLSI will not adversely impact any race/ethnic or gender group if used for screening candidates for recruiting duty. However, only a limited number of race/ethnic groups were examined. Data from larger samples of job applicants, representing a broader range of race/ethnic groups, are needed to draw more definitive conclusions regarding the effects of screening by applicants' gender and race/ethnicity.

Table 16. NLSI Scores for Demographic Subgroups

		Mean NLSI Score	Standard Deviation	N
Gender	Male	49.83	8.56	3,654
	Female	50.75	8.09	348
Race/Ethnicity	Black	49.73	8.61	1,020
	Caucasian	49.64	8.54	2,322
	Hispanic	51.46	8.07	425
	American Indian/Alaskan Native	52.31	8.18	20
	Asian/Pacific Islander	49.54	8.81	93
	Other/Unknown	53.19	7.69	42

Comparison of Paper-and-Pencil vs. Online Format NLSI

Of the 4,586 recruiters for whom NLSI data were available, 2,143 took the paper-and-pencil version, and 2,443 took the online version. The means and standard deviation scores on the empirically-keyed total NLSI score were very similar (paper-and-pencil mean = 49.06, standard deviation = 8.43; online mean = 50.89, standard deviation = 8.05). Accordingly, we conducted the validation analyses with data from the two versions combined.

28

NLSI and Attrition from Recruiter Training

In addition to recruiter production and performance ratings, we also examined the relationship between the NLSI and attrition from training in the ARC. Students attrit from the ARC for various reasons. A majority of the reasons for attrition are performance-related (see Figure 2 below).

Legend:
- ■ Performance-related (Academic, Disciplinary/Misconduct, Motivational)
- □ Non Performance-related (Emergency Leave, Compassionate Leave, Did not meet Pre-reqs)
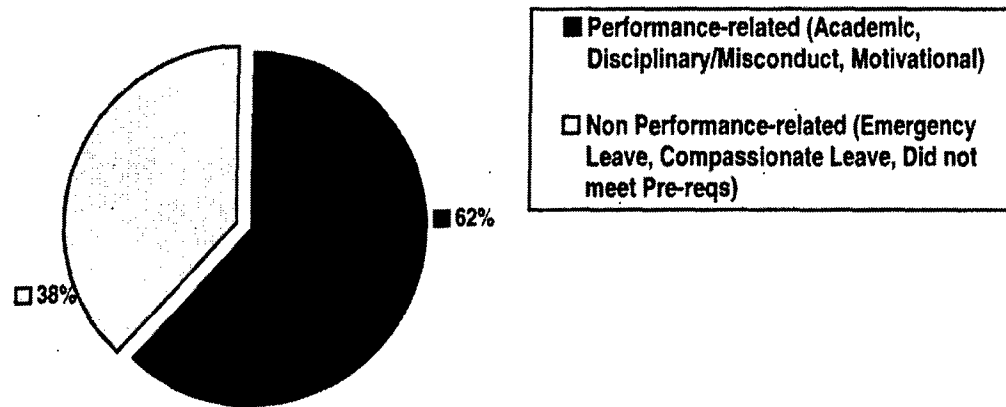
Figure 2. Reasons for ARC Attrition.

The NLSI total score predicted attrition from the recruiter training in the ARC ($r = -.10$, $N = 3,714$). Recruiters scoring in the bottom 5% on the NLSI had a 29.2% total attrition rate compared to a 9.8% total attrition rate for recruiters scoring in the top 95% on the NLSI (see Figure 3). The same pattern is evident for the performance-related attrition. The correlations between individual NLSI scales and attrition from recruiter training ranged from .00 to .11 (absolute values), with a median correlation of .07 (absolute value; Ns range from 1,457-3,240).
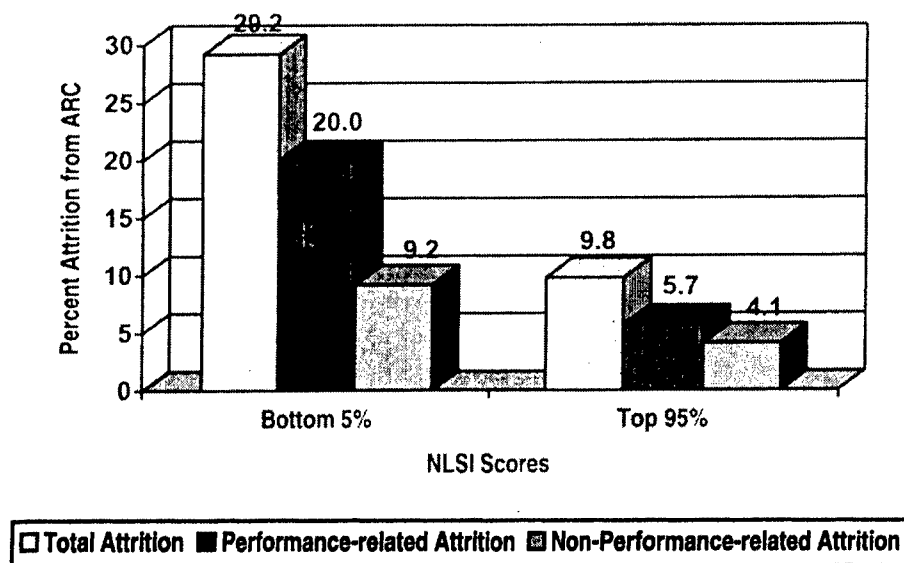
Figure 3. NLSI Scores Predict ARC Attrition.

Recruiting Duty Relief

Even after considerable training, USAREC will sometimes determine that a new recruiter is ineffective, will relieve the recruiter from duty, and return the Soldier to his or her previous MOS. Accordingly, recruiting duty relief seemed to be a potential criterion measure. To determine the viability of such a measure, we examined the recruiter production database for the numbers of recruiters in the relieved category.

Approximately 2,500 new recruiters were tested on the NLSI in 2002. Of these, recruiting duty relief information was available for 2,393 recruiters. At the time of analysis (July 2003), 2,343 of these were still recruiters, while 50 had been relieved from recruiting duty.

Upon inspection of the reasons for relief, two primary categories surfaced: "relieved" and "relieved without prejudice." Those recruiters in the "relieved" category (N = 32) were relieved from recruiting duty due to poor recruiter performance. In contrast, those recruiters included in the "relieved without prejudice category" (N = 18) were relieved from recruiting duty for reasons unrelated to their performance. These recruiters were classified as unqualified (UNQ) because of medical, financial, or spousal reasons. Because the main objective of the NLSI is to predict recruiter performance, we were mainly concerned with the recruiters in the "relieved" category. Unfortunately, the number of relieved cases was so small in our sample that it was problematic as a criterion measure and additional validity analyses were not conducted at this time. However, we continue to follow the recruiters tested on the NLSI. As the sample matures and as more data becomes available, we plan to investigate the relationships between the NLSI and this criterion measure.

NLSI Part III - SJT Scoring

We used several approaches to develop a key for the SJT portion of the NLSI. Scoring keys can be developed rationally, based on effectiveness ratings from subject matter experts, or empirically, by comparing answers on the SJT with some external criterion measure (e.g., performance ratings). We did not have a sample of expert recruiters to help develop a scoring key for the revised SJT, therefore we used an empirical keying approach. Research has indicated that empirical keying can be just as effective as keying based on subject matter experts' ratings (MacLane, Barton, Holloway-Lundy, & Nickels, 2001; Paullin & Hanson, 2001; Weekly & Jones, 1999).

Attempting to empirically key the SJT against recruiter performance measures (i.e., recruiter performance ratings and recruiter production), we explored a number of empirical keying approaches based on their success in previous research. One empirical keying technique is the Correlational Method. In this method, the sample is divided into two contrasting groups (e.g., high and low performing recruiters), and the response options are dichotomously scored (e.g., selected or not selected). A phi correlation is computed, characterizing the relationship between the contrasting group variable and the dichotomous response variable. This correlation is used to assign weights to the response options such that a unit weight is assigned only to statistically significant relationships. Other weights can also be assigned based upon the magnitude and direction of the zero-order correlation. Generally, research exploring different empirical keying strategies has found this strategy produced scores that were significantly related to performance in cross-validation samples (Krokos, Meade, Cantwell, Pond, & Wilson, 2004; Paullin & Hanson, 2001).

Another technique is the Vertical Percent Method. Similar to the Correlational Method, the procedure begins with the formation of contrasting groups based on criterion scores. Next, the percentage of persons choosing each response option and the difference between percentages from these two groups are calculated. Values of weights may be determined in a variety of ways, including using absolute differences. Using the absolute difference approach, the actual difference in percentages is used as the weight. Larger differences enable us to predict more reliably who should be classified into each group, thus larger weights are assigned to these response options. Positive weights are given to options that the high performance group selects more often than the low performance group. Conversely, negative weights are assigned to options chosen by a larger proportion of the low performance group. Using this method, all response options that differentiate between the two groups receive a non-zero weight. The Vertical Percent Method has shown similar or slightly higher validity when compared to other keying methods in some studies (Devlin, Abrahams, & Edwards, 1992; Paullin & Hanson, 2001). However, Krokos et al. (2004) found the Correlational Method to have superior validity using a small student sample.

We first used the Vertical Percent, Absolute Difference Method to develop an empirical key. After splitting the sample of participants into development and cross-validation sub-samples, we used the development sub-sample data to derive empirical keys for the prediction of performance ratings and recruiter production. These empirical keys were then used to produce scores for the cross-validation sub-sample. However, these empirically keyed scores did not exhibit significant validity for the prediction of performance ratings or recruiter production. Other empirical keying methods were utilized in addition to the Absolute Difference Method, but as with the previous method, the items failed to cross-validate.

As we were unable to develop a key that successfully cross-validated, we did not conduct further analyses with the SJT portion of the NLSI. In summary, the SJT items did not add value to the NLSI for recruiter screening purposes. However, the RRS has expressed interest in using the Army Recruiter Situational Judgment Test developed in the concurrent validation for training purposes at the school.

Summary

In sum, these results indicate NLSI scores predict recruiter production. Moreover, results presented here may be an underestimate of the true relationship between the NLSI and recruiter performance. The realities of the recruiting environment add challenges to the prediction of individual recruiter production. For example, we calculated an individual production average for each recruiter in our sample. However, one of the goals of station missioning is to move away from strictly individual recruiting goals toward more flexible mission requirements at the station-level. This is quite effective at the station level, but increases imprecision in the measurement of individual production, as the effort to recruit is shared among recruiters in a station.

In addition to recruiter production, NLSI scores also predicted attrition from recruiter training. The NLSI demonstrated modest, statistically significant correlations with these two important criterion measures. The NLSI did not significantly correlate with the performance rating criterion. However, these results are not surprising given the small sample size and the emphasis placed on production in the development of the empirical key.

Chapter 7 - Conclusion

The work described in this technical report constitutes a multi-year program to implement proctored, online testing and to conduct research on the prediction of Army recruiter performance. Since the implementation of the online test, thousands of recruiters have taken the NLSI from locations around the world. The validation results indicate the NLSI predicts recruiter success, both in training and on-the job.

As a result of this research, there is a working system to deliver secure, proctored testing in Army DTFs. A number of Army and contractor organizations worked together to plan, test, and implement a secure system to share data to notify and schedule Soldiers for testing, and deliver proctored testing across the world. In addition, procedures were developed to securely share results among RRS decision-makers, HRC, and researchers conducting the validation research. Finally, the system was designed to accommodate changes to the scoring procedures or to the items themselves. Thus, future work to further refine the scoring key or add new items can be implemented quickly.

After several years of NLSI testing and criterion measure development and collection, we have developed the NLSI Predictive Validation Database. This large database contains demographic and background information, NLSI scores, training outcome measures, and measures of on-the-job performance (i.e., recruiter detail relief, production indices, and performance ratings). This database can serve as the basis for future work to refine the NLSI items and the scoring key. We will continue to investigate the relationships between the NLSI and several criterion measures as the sample of recruiters mature in the job. Many of the recruiters in the research presented here had less than one year of recruiting experience, and we will continue to follow those recruiters through their second and third year of recruiting detail.

Using this large pool of NLSI and criterion data, we refined the NLSI scoring key that was originally developed in the concurrent validation research. The results of the predictive validation research have demonstrated that the NLSI can be a valuable tool for screening Soldiers for recruiting duty. As recruiting becomes more and more difficult, individual differences in recruiters' aptitude and skills may become even more critical to successful performance.

Selecting recruiters with higher NLSI scores will result in higher levels of production, and other important benefits such as increased levels of job satisfaction, lower levels of stress, and higher quality of life may result from selecting those Soldiers best suited for recruiting duty. In the concurrent

33

validation research, recruiters with higher NLSI scores reported experiencing higher quality of life compared to recruiters with lower NLSI scores (White, Kubisiak, Horgen & Young, 2004). Selecting Soldiers who will likely succeed in recruiting will also benefit station commanders, as they will spend less time counseling poor performing recruiters. In addition, Soldiers who are less effective recruiters may be highly effective in another MOS and provide a more valuable contribution to the Army in another function. Finally, in addition to a number of benefits to be gained by screening out Soldiers, the cost of the screening itself is nominal. ARI, USAREC, and PDRI broke new ground to implement recruiter testing in proctored settings worldwide, resulting in considerable efficiencies over the traditional paper-and-pencil testing technologies.

The NLSI validation data supports an initial use of the NLSI for screening a small percentage (e.g., 5%) of Soldiers likely a poor fit for recruiting duty. A computerized version of the NLSI for testing at DTFs worldwide was developed to support assessment of large samples of Soldiers needed to populate the database of potential candidates for recruiting duty and possibly other specialties. In addition, other research from a preliminary, small sample validation indicates that the NLSI is related to Drill Sergeant success in mentoring and training, as measured by performance ratings (Kubisiak, et al., 2005). Future research is needed to guide the potential refinement of the NLSI as a classification tool for multiple Army NCO positions.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Borman, W. C., Campbell, C. H., Hanson, M. A., Hedge, J. W., Knapp, D. J., McGuire, D. P., Pulakos, E. D., & Whetzel, D. (1990). Development of scores for second-tour performance measures. In J. P. Campbell & L. M. Zook (Eds.), *Building and Retraining the Career Force: New Procedures for Accessing and Assigning Army Enlisted Personnel, Annual Report, 1990 Fiscal Year* (pp. 247-336). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Borman, W. C., Horgen, K. E., Birkeland, S. A., Penney, L. M., Sutton, M. J., & Mills, L. J. (2001). *Development of recruiter assessment measures for the U.S. Army* (Institute Report #382). Tampa, FL: Personnel Decisions Research Institutes, Inc.

Borman, W. C., Horgen, K. E., & Penney, L. M. (2000). *Overview of ARI recruiting research* (ARI Research Note 2000-07). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Borman, W. C., Penney, L. M., Horgen, K. E., Birkeland, S. A., Hedge J. W., Bowles, S., & Held, J. D. (2001, April). *Recruiting research in the U.S. Army and Navy*. Invited paper presented at the International Workshop on Military Recruitment and Retention in the 21st Century, The Hague, NL.

Borman, W. C., Rosse, R. L., & Toquam, J. L. (1982). *The impact of environmental factors and considerations of recruit quality on Navy recruiter production* (Institute Report #78). Minneapolis, MN: Personnel Decisions Research Institutes.

Borman, W. C., White, L. A., Bowles, S. V., Horgen, K. E., Kubisiak, U. C., & Penney, L. M. (2003, November). *Update on recruiting research in the U.S. Army*. Paper presented at the 45th Annual Conference of the International Military Testing Association, Pensacola, Florida.

Borman, W. C., White, L. A., Young, M. C., Kubisiak, U. C., Horgen, K. E., Penney, L. M., Lee, V. B., & Connell, P. W. (2004, November). *U.S. Army recruiter selection research: Development of the Noncommissioned Officer Leadership Skills Inventory (NLSI)*. Paper presented at the 46th Annual Conference of the International Military Testing Association, Brussels, Belgium.

Bowles, S. V., Kubisiak, U. C., Horgen, K. E., White, L. A., Young, M. C., & Borman, W. C. (2003, May). *Identifying High Potential Recruiter Candidates*. Symposium conducted at the U. S. Army Accessions Command, Accessions Research Consortium, Hampton, VA.

Devlin, S. E., Abrahams, N. M., & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. *Military Psychology, 4*, 119-136.

Goldberg, L., Kimko, D., & Lewis, L. (2005, October). *Applications of the Army enlistment early warning system: Forecasts in FY06*. Symposium conducted at the U. S. Army Accessions Command, Accessions Research Consortium, Hampton, VA.

Hanson, M. A., & Borman, W. C. (1995, May). *Construct validation of a measure of supervisory job knowledge*. Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, Orlando, FL.

Horgen, K. E., Penney, L. A., Birkeland, S. A., Kubisiak, U. C., & Borman, W. C. (2001, August). *Army recruiter situational judgment test*. In M. Rumsey (Chair), New directions in recruiter assessment and selection. Symposium conducted at the 109th Annual Convention of the American Psychological Association, San Francisco, CA.

Kilcullen, R. N., Chen, G., Zazanis, M. M., Carpenter, T., & Goodwin, G. (1999). *Adaptable performance in unstructured environments*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Atlanta, GA.

Kilcullen, R. N., Mael, F. A., Goodwin, G. F., & Zazanis, M. M. (1999). Predicting U.S. Army Special Forces field performance. *Journal of Human Performance in Extreme Environments, 4*, 53-63.

Kilcullen, R. N., White, L. A., Zacarro, S., & Parker, C. (2000). *Predicting Managerial and Executive Performance*. Paper presented at the 15th Annual Conference, Society for Industrial and Organizational Psychology, New Orleans, LA.

Krokos, K. J., Meade, A. W., Cantwell, A. R., Pond, S. B., & Wilson, M. A. (2004). *Empirical Keying of Situational Judgment Tests: Rationale and Some Examples*. Poster presented at the 19th Annual Conference, Society for Industrial and Organizational Psychology, Chicago, IL.

Kubisiak, U. C., Horgen, K. E., Connell, P. W., Lentz, E., Xu, X., Borman, W. C., White, L. A., & Young, M. C. (2005). *Concurrent validation of the NLSI for U.S. Army Drill Sergeants* (Institute Report #499). Tampa, FL: Personnel Decisions Research Institutes.

Kubisiak, U. C., McGonigle, T., Horgen, K. E., Borman, W. C., Kaufman, J. D., & Casper, W. J. (2003). *Identifying and evaluating USAREC station-level business practices: Literature review* (Institute Report #426). Tampa, FL: Personnel Decisions Research Institutes.

MacLane, C. N., Barton, M. B., Holloway-Lundy, A. E., & Nickels, B. J. (2001). *Keeping score: Empirical vs. expert weights on situational judgment responses.* Paper presented at the 16th annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Paullin, C. J., & Hanson, M. A. (2001). *Comparing the validity of rationally derived and empirically derived scoring keys for a situational judgment inventory.* Poster presented at the 16th Annual Conference, Society for Industrial and Organizational Psychology, San Diego, CA.

Penney, L. M. (2004). *Guide for creating the Army recruiter production database.* (unpublished Institute Report). Tampa, FL: Personnel Decisions Research Institutes, Inc.

Penney, L. M., Horgen, K. E., & Borman, W. C. (1999). *An annotated bibliography of recruiting research conducted by the U. S. Army Research Institute for the Behavioral and Social Sciences* (Institute Report #340). Tampa, FL: Personnel Decisions Research Institutes.

Penney, L. A, Horgen, K. E., Kubisiak, U. C., Borman, W. C., & Birkeland, S. (2002, August). Measurement of recruiter performance and productivity. In M. Rumsey (Chair), *New finding on prediction of recruiting success.* Symposium conducted at the 110th Annual Convention of the American Psychological Association, Chicago, IL.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Sutton, M. J., Horgen, K. E., Borman, W. C., & Kubisiak, U. C. (2001). *Exploring new screening processes for U. S. Army recruiters* (Institute Report #375). Tampa, FL: Personnel Decisions Research Institutes, Inc.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures (4th ed.).* Bowling Green, OH: Author.

Uniform guidelines on employee selection procedures, 43 *Federal Register 38290-38315* (1978).

United States Army Recruiting Command (USAREC). (2005). Retrieved August 28, 2005, from http://www.usarec.army.mil/info.html

Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52,* 679-700.

White, L. A., Borman, W. C., & Bowles, S. (2001, August). *New tools for selecting high potential Army recruiters.* Paper presented at the 109th Annual Convention of the American Psychological Association. San Francisco, CA.

White, L. A., Borman, W. C., Penney, L. A., Kubisiak, U. C., Horgen, K. E., Bowles, S. A. & Mills, L. J. (2002, August). Concurrent validation of new recruiting screening measures for Army recruiters. In M. Rumsey (Chair), *New findings on prediction of recruiting success.* Symposium conducted at the 110th Annual Convention of the American Psychological Association, Chicago, IL.

White, L. A., Kubisiak, U. C., Horgen, K. E., & Yound, M. C. (2004, October). *Overview of NLSI Research Program.* Briefing presented to COL Jack Collins, Commandant RRS, Columbia, SC.

White, L. A., & Young, M. C. (2001, April). *Validation of a faking-resistant measurement of temperament constructs.* Paper presented at the 16th Annual Conference, Society for Industrial and Organizational Psychology, San Diego, CA.

Young, M. C., Heggestad, E. D., Rumsey, M. G., & White, L. A. (2000). *Army pre-implementation research findings on the assessment of individual motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Young, M.C., McCloy, R.A., Waters, B.K., & White, L.A. (2004). Introduction: An overview of AIM and preliminary efforts to support its operational use. In Knapp, D.J., Heggestad, E.D., & Young, M.C. (Eds.), *Understanding and Improving the Assessment of Individual Motivation (AIM) in the Army's GED Plus Program.* (ARI Study Note 2004-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Young, M.C., White, L.A., Heggestad, E.D., & Barnes, J.D. (2004, July). *Operational validation of the Army's new pre-enlistment screening measure.* Paper presented at the annual meeting of the American Psychological Association, Honolulu, HI.

Appendix A – NLSI Part I Scales and Definitions

Tolerance for Ambiguity

This scale measures a person's preference for work environments in which the problems (and potential solutions) are unstructured and ill-defined. Those with high tolerance for ambiguity are comfortable working in rapidly changing work environments. Individuals scoring low prefer highly structured and predictable work settings.

Hostility to Authority

The degree to which a person respects and is willing to follow legitimate authority figures. High scorers are expressively angered by authority figures and may actively disregard their instructions and policies. Low scorers accept directives from superiors and easily adapt to structured work environments.

Social Perceptiveness

This scale measures the degree to which a person can discern and recognize others' emotions and likely behaviors in interpersonal situations. Persons high in social insight are good at understanding others' motives and are less likely to be "caught off guard" by unexpected interpersonal behaviors.

Interpersonal Skill

This scale measures the degree to which a person establishes smooth and effective interpersonal relationships with others. Interpersonally skilled individuals are good listeners, behave diplomatically, and get along well with others. Persons with low scores on this measure have difficulty working with others and may intentionally or unconsciously promote interpersonal conflict and cause hurt feelings.

Emergent Leadership

The scale measures the degree to which a person takes on leadership roles in groups and in his or her interactions with others. High scorers on this scale are looked to for direction and guidance when group decisions are made and readily take on leadership roles.

Conscientiousness

This scale measures the degree to which a person is achievement-oriented and dedicated to work. Persons high in conscientiousness are hard working, persistent, self-disciplined, and deliberate. Individuals scoring low are more careless in work-related activities, prefer leisure activities to work, and can be easily distracted from work-related tasks.

Self-Esteem

This scale measures the degree to which a person feels good about oneself as a person and has confidence in one's own abilities. Individuals with high self-esteem feel successful in past undertakings and expect this to continue in the future. Low scorers have feelings of personal inadequacy, lower self-efficacy, and lack confidence in their ability to be successful.

Empathy

This scale measures the degree to which a person understands and shares others' thoughts and emotions. High scorers are sensitive, and find it difficult to watch the suffering of others.

Appendix B – NLSI Part II Scales and Definitions

Work Motivation

The tendency to strive for excellence in the completion of work-related tasks. Persons high on this construct seek challenging work activities and set high standards for themselves. They consistently work hard to meet these high standards.

Adjustment

The tendency to have a uniformly positive affect. Persons high on this construct maintain a positive outlook on life, are free of excessive fears and worries, and have a feeling of self-control. They maintain their positive affect and self-control even when faced with stressful circumstances.

Agreeableness

The tendency to interact with others in a pleasant manner. Persons high on this construct get along and work well with others. They show kindness, while avoiding arguments and negative emotional outbursts directed at others.

Dependability (Non-delinquency)

The tendency to respect and obey rules, regulations, and authority figures. Persons high on this construct are more likely to stay out of trouble in the workplace and avoid getting into difficulties with law enforcement officials.

Leadership (Dominance)

The tendency to seek out and enjoy being in leadership positions. Persons high on this scale are confident of their abilities and gravitate towards leadership roles in groups. They feel comfortable directing the activities of other people and are looked to for direction when group decisions have to be made.

Physical Conditioning

The tendency to seek out and participate in physically demanding activities. Persons high on this construct routinely participate in vigorous sports or exercise, and enjoy hard physical work.

Appendix C – Development of a Production Quality Index

Development of a Production Quality Index

The primary criterion measure for NLSI validation efforts is recruiters' raw production (e.g., signed contracts). Although raw production provides appropriate information regarding the number of prospects a recruiter is able to bring in, it provides no information about the quality of these recruits. Production quality is an important criterion to consider for the Army, as Congress mandates yearly accession goals for various recruit-quality levels. Our goal was to determine the feasibility of developing an index of production quality and to evaluate the reliability of this new measure.

Defining Quality

In general, the Army defines production quality based, in part, on the Armed Forces Qualification Test (AFQT) scores. The AFQT is administered to every new recruit before they access and is used to determine an individual's eligibility to enlist. The AFQT is a combination of scores from four subtests included in the ASVAB (Arithmetic Reasoning, Mathematics Knowledge, Paragraph Comprehension, and Word Knowledge). Each AFQT score is expressed as a percentile in reference to youth population norms. For the purposes of these analyses, we define high quality as at or above the mean AFQT score (categories I-IIIA).

Creating an AFQT Quality Index

The first step in the creation of an AFQT-based quality index was to compile information regarding recruit quality and recruiter production. With the help of USAREC and ARI, production information from October, 2000, to May, 2003 was obtained on over 12,000 recruiters. Note that the data for these analyses were based on the population of USAREC recruiters, not the sample of recruiters used in the predictive validity analyses. This information included both the gross and net number of monthly I-IIIA high quality contracts brought in by each recruiter.

Next, we calculated the monthly gross and net percentage of I-IIIA contracts brought in by each recruiter. These values were computed by dividing the gross (net) number of I-IIIA contracts by the overall gross (net) number of contracts for each month. Thus, for each recruiter, we had an index of quality that reflected the percentage of high quality contracts (both gross and net) brought in by that recruiter for each of the months that they were considered on production. To simplify further analyses, we used signed contracts, or gross number of contracts.

The next issue involved correcting for environmental factors that may impact recruiter production. A recruiter's production may be influenced by territorial factors that are beyond his or her control (Borman, Rosse, and Toquam, 1982). Therefore we adjusted our gross production quality index for territorial

49

differences using the mean quality estimates for the target territories. The reason for using a territorial adjustment here and not doing so with the production criterion measures is that it seemed likely there would be stronger territory effects for recruit quality than for general availability of recruits.

In addition to territorial influences, seasonal differences may also significantly affect recruiter production quality (Penney et al., 2002). For example, it is probably easier to recruit high quality contracts during the summer before high school begins as opposed to late in the spring semester. Most high quality students have made future plans by this point in time (e.g., attending college). Thus, our gross quality index was also adjusted for seasonal influences on recruiting at the brigade, battalion, and company level using the mean quality estimates for the targeted months.

Reliabilities

Before the reliabilities associated with this new index could be calculated, we had to decide on the number of months necessary to employ. As mentioned, we obtained data for 32 months, but not all recruiters were on production for all 32 months. On average, we had 18.4 months of production data for the recruiters in our sample. Thus, the final quality score for some recruiters was based on as few as 3-4 months of data, whereas a small number had as many as 32 months. Because the stability of production quality is likely to be higher when more months' data are averaged, including scores on only a few months of data may attenuate the observed relationships with other variables. We therefore examined the reliability of production quality scores based on a varying number of months (see Table 17).

Table 17. Reliability of Production Quality Scores

Number of Months of Production Information	Quality (% of I-IIIA contracts corrected for month and brigade)	Quality (% of I-IIIA contracts corrected for month and battalion)	Quality (% of I-IIIA contracts corrected for month and company)	Individual Recruiter Production (Average contracts corrected for month and brigade)
12 months	.69	.68	.65	.81
11 months	.67	.66	.64	.80
10 months	.64	.63	.61	.77
9 months	.60	.60	.58	.75
8 months	.56	.56	.54	.72
7 months	.52	.51	.49	.69
6 months	.45	.44	.42	.66
5 months	.40	.39	.37	.63
4 months	.33	.32	.30	.60
3 months	.26	.26	.24	.49
2 months	.19	.19	.17	.38

Twelve months was used as an upper limit. As expected, the reliability of production quality scores increases as the number of months employed increases. A reasonable cut-off is 9 months; with fewer than nine months of data, the reliability of the production quality index appears to be questionable. Also, we decided to use battalion level to correct for territory. The reasoning was that battalion and brigade provide very similar levels of reliability and homogeneity of environmental factors seems of primary importance in this context.

Recruit Quality and Production

We were also interested in the relationship between gross production and recruit quality. That is, we wished to determine whether recruiters who bring in a large number of contracts each month tended to bring in a higher percentage of high quality contracts (positive relationship), more low quality contracts (negative relationship), or whether the number of contracts brought in by an individual recruiter had no bearing on the quality of his or her recruits (no relationship).

For these analyses, we computed a production quality index. That is, for each recruiter, we computed the mean gross percentage of A contracts brought in by that recruiter across 32 months of production. We also created a similar index for gross production. Consistent with earlier findings, both of these indices were corrected for both month and battalion.

Overall, we found that recruiters' adjusted monthly gross production had a significant, negative relationship with the recruit quality ($r = -.15$, $p < .01$, $N=11,612$). In other words, recruiters who bring in large numbers of recruits each month also tend to enlist larger numbers of lower quality recruits. This is in comparison to recruiters who bring in small numbers of recruits each month, but whose recruits tend to be of higher quality. Apparently, high production recruiters tend to recruit a larger percentage of low quality prospects compared to their lower producing counterparts.

Although these results suggest that is it possible to create a reliable quality index, this effort should be interpreted as a preliminary investigation of recruit quality. The recruit quality index developed here does not take into account all criteria for quality considered by the Army. The most obvious is recruit educational attainment. Educational status is taken into account by recruiters and weighted according to a number of subcategories ranging from college graduate to non-high school graduate. Building on the promising foundation of the current research, future research should incorporate factors such as education status into a production quality index. Because of the preliminary status of the quality index, we did not use these criteria in subsequent validation research with the NLSI.

Appendix D - Army Recruiter Performance Rating Scales

Army Recruiter Performance Rating Scales

54

A. Locating And Contacting Qualified Prospects

Contacting large numbers of persons likely to join the Army; skillfully using available recruiting aids to gain the attention and interest of young persons eligible for Army service; knowing where and when to prospect; persisting in prospecting and following up on leads even under considerable adversity.

Needs Considerable Improvement	Needs Improvement	Effective	Very Effective
Exerts little effort prospecting; for example, often fails to follow up on even promising leads, and uses recruiting tools (e.g., telephone, RTools) sparingly and ineffectively.	Exerts effort prospecting, but may use a limited number of recruiting tools and may not spend enough time with or direct sufficient effort toward the most productive sources and prospects.	Uses a number of sources and recruiting tools for prospecting and is effective at locating and contacting qualified prospects.	Displays exceptional ingenuity and energy and uses a wide variety of recruiting tools very effectively to locate and contact qualified prospects.
❶ ❷	❸ ❹ ❺	❻ ❼ ❽	❾ ❿

55

B. Gaining And Maintaining Rapport

Being hospitable to prospects in the office; gaining the trust and respect of prospects; adjusting to applicants' styles and acting appropriately with different types of applicants.

Needs Considerable Improvement		Needs Improvement		Effective		Very Effective	
Is very poor at gaining and maintaining rapport; appears disinterested in applicants or may answer questions in an impersonal way.		Has trouble interacting with certain prospects; sometimes appears disinterested in a prospect or may have a standard approach to interacting that is inappropriate for some prospects.		Is typically able to put prospects at ease, and maintains good rapport with them; interacts with most prospects in a warm and friendly way.		Interacts very effectively with all types of prospects; is excellent at gaining and maintaining rapport and establishing trust with prospects.	
❶	❷	❸	❹	❺	❻	❼	❽	❾	❿

C. Obtaining Information From Prospects And Making Good Person-Army Fits

Demonstrating good listening and interviewing skills; making accurate judgments and suggesting options to match prospects' needs and preferences; effectively obtaining information about prospects from other sources (e.g., high school principal, parents) to assess their qualifications and needs.

Needs Considerable Improvement			Needs Improvement			Effective			Very Effective	
Is very poor at getting prospects to reveal their needs and buying motives, making it difficult to suggest appropriate Army opportunities; may suggest features or programs that clearly don't interest prospect.			Sometimes fails to learn enough about prospects to identify their primary needs and buying motives; may suggest Army features and benefits that do not result in a good match with the individual's needs.			Is good at blueprinting most prospects, evaluating their needs, and then discussing Army opportunities appropriate for meeting those needs.			Always blueprints effectively, identifying prospects' needs and career motivations and then is excellent at emphasizing Army features and opportunities that address these needs and motives.	
❶	❷	❸	❹	❺	❻	❼	❽	❾	❿	

57

D. Salesmanship Skills

Skillfully persuading prospects to join the Army; selecting and adapting selling techniques appropriate to different prospects; effectively presenting Army benefits and opportunities; identifying and overcoming objections to joining the Army; persisting to close the sale.

Needs Considerable Improvement		Needs Improvement		Effective		Very Effective	
Fails to present Army features/benefits to influence individual prospects and is frequently unable to identify or overcome objections to joining the Army; often misses opportunities to close even when it's clearly appropriate to do so.		Presents Army features and benefits in a way that is sometimes not suitable for an individual prospect and may not identify prospects' objections to the Army; at times, misses opportunities to close.		Presents Army features/benefits so that most prospects become more interested in the Army; recognizes and is prepared to overcome frequently heard objections about the Army; knows when and how to close in most situations.		Presents Army life and benefits in a highly appropriate and convincing way for each prospect, and is very adept at identifying and overcoming any objections about the Army; never misses opportunities to close.	
❶	❷	❸	❹	❺	❻	❼	❽	❾	❿

E. DEP/DTP Maintenance

Skillfully relaying accurate information about BCT, Army life, and the Army's expectations so that prospects/recruits know what to expect; training DEP/DTPs to standards on pre-BCT tasks (e.g. hand salute, physical fitness standards); following up and maintaining relationships with DEP/DTPs and their families, effectively overcoming objections and buyer's remorse to ensure DEP/DTPs maintain enlistment intentions.

Needs Considerable Improvement	Needs Improvement	Effective	Very Effective
Fails to maintain contact with DEP/DTPs after they enlist; provides DEP/DTPs with minimal or inaccurate information about Army life; rarely prepares them for BCT, and exerts little effort to counsel individuals who no longer wish to enlist.	Sporadically contacts DEP/DTPs after mandatory follow-up; may miss signs of buyer's remorse; spends little time training DEP/DTPs to standard on pre-BCT tasks, and may provide incomplete information about Army life.	Follows up with DEP/DTPs and their families as needed; responds to concerns in a sensitive manner; works to train DEP/DTPs to standard and accurately describes Army expectations and BCT to prepare them for Army life.	Maintains contact with and provides emotional support to all DEP/DTPs and their families; thoroughly prepares DEP/DTPs for BCT and Army life by training them to standard and providing complete, detailed information about Army life and expectations.
❶ ❷	❸ ❹ ❺	❻ ❼ ❽	❾ ❿

59

F. Establishing And Maintaining Good Relationships In The Community

Contacting and working effectively with individuals and agencies capable of helping with prospects; presenting a good image and building a good reputation for the Army by developing positive relationships with persons in the community; presenting a good Army image through appearance, language, and demeanor.

Needs Considerable Improvement	Needs Improvement	Effective	Very Effective
Avoids personal contact or alienates individuals in the community by making demands or failing to honor commitments; presents negative image of the Army by poor personal appearance or behavior.	Does not make regular contact with community agencies that might be helpful in recruiting, and does not develop relationships fully; is not particularly alert to opportunities to promote the Army.	Spends productive time with individuals/agencies, and keeps them informed of most Army activities; may arrange Army activities for community persons who can help in recruiting.	Is exceptionally alert to and adept at developing relationships with relevant individuals and community agencies, and promotes a strongly positive image of the Army; may volunteer off-duty time to help in the community.
❶ ❷	❸ ❹ ❺	❻ ❼ ❽	❾ ❿

G. Organizing Skills/Time Management

Planning ahead; organizing time efficiently; completing paperwork accurately and on time; keeping track of appointments; not wasting time; knowing how to locate important reference materials (e.g. qualifications, regulations, standards); ensuring that recruits are processed by quickly and efficiently getting them to the MEPS and into the Army.

Needs Considerable Improvement	Needs Improvement	Effective	Very Effective
Consistently fails to complete necessary forms or may use wrong forms; organizes time poorly and does not maintain applicant log/planner.	Sometimes completes paperwork late, occasionally with significant errors; is somewhat inefficient in use of time and may at times schedule appointments without considering other events.	Usually completes paperwork on time and with few errors; keeps an accurate applicant/log planner and generally uses time efficiently.	Accurately completes all paperwork, prior to or as scheduled; devises plans so as to achieve own and station goals; maintains complete calendar of relevant events and schedules work activities very efficiently and effectively.
❶　　❷	❸　　❹　　❺	❻　　❼　　❽	❾　　❿

H. Supporting Other Recruiters And USAREC

Coordinating activities with and supporting other recruiters to maximize the productivity of the station and battalion; complying with and supporting orders and directives from higher levels; mentoring or providing helpful tips and constructive feedback to other Army recruiters, especially if they are new.

Needs Considerable Improvement	Needs Improvement	Effective	Very Effective
Rarely cooperates with, supports, or helps other recruiters, even if requested, and lets others carry the recruiting load; may interfere with the group effort by withholding important information or not coordinating own activities with others.	May assist other recruiters when specifically asked but does not look for opportunities to help or support others; often complains about mission or having to work extra time.	Supports the Command in ways that are helpful; usually places station/battalion mission above personal goals, and generally cooperates and works well with fellow recruiters.	Is always enthusiastic and works to build group spirit; consistently helps other recruiters, even when he/she is busy; always shares information so as to increase group production.

❶   ❷   ❸   ❹   ❺   ❻   ❼   ❽   ❾   ❿

62